2016

Short-term Building Energy Model Recommendation System: A Meta-learning Approach

Can Cui
Arizona State University

Teresa Wu
Arizona State University

Mengqi Hu
The University of Illinois at Chicago

Jeffery D. Weir
Air Force Institute of Technology

Xiwang Li
Arizona State University

Recommended Citation

Cui, C., Wu, T., Hu, M., Weir, J. D., & Li, X. (2016). Short-term building energy model recommendation system: A meta-learning approach. Applied Energy, 172(15 June), 251–263. https://doi.org/10.1016/j.apenergy.2016.03.112

Short-TermBuildingEnergyModel RecommendationSystem: A Meta-Learning Approach

Can Cui[a]
School of computing, informatics, and decision systems engineering
Arizona State University
699 S. Mill Ave., Tempe, AZ 85281, USA
ccan1@asu.edu
Teresa Wu[a] (Corresponding author)
Room 308, 699 S. Mill Ave., Tempe, AZ 85281, USA
Teresa.Wu@asu.edu, 1-(480)-965-4157
Mengqi Hu[b]
Department of Mechanical and IndustrialEngineering
University of Illinois at Chicago
842 W. Taylor St., Chicago, IL 60607
mhu@uic.edu
Jeffery D. Weir[c]
Department of operational sciences
Air Force Institute of Technology
2950 Hobson Way, Wright-Patterson Afb, Ohio 45433, USA
Jeffery.Weir@afit.edu
Xiwang Li[d]
Center for Green Buildings and Cities, Graduate School of Design
Harvard University
20 Sumner Rd, Cambridge, MA, 02138
xiwang_li@gsd.harvard.edu

Abstract

High-fidelity and computationally efficient energy forecasting models for building systems are needed to ensure optimal automatic operation, reduce energy consumption, and improve the building's resilience capability to power disturbances. Various models have been developed to forecast building energy consumption. However, given buildings have different characteristics and operating conditions, model performance varies. Existing research has mainly taken a trial-and-error approach by developing multiple models and identifying the best performer for a specific building, or presumedone universal model form which is applied on different building cases. To the best of our knowledge, there does not exist a generalized system framework which can recommend appropriate models to forecast the building energy profiles based on building characteristics. To bridge this research gap, we propose a meta-learning based framework, termed Building Energy Model Recommendation System (BEMR). Based on the building's physical features as well as statistical and time series meta-features extracted from the operational data and energy consumption data, BEMR is able to identify the most appropriate load forecasting model for each unique building.Three sets of experiments on 48 test buildings and one real building are conducted. The first experiment is to test the accuracy of BEMR when the training data and testing data cover the same condition. BEMR correctly identified the best modelon 90% of the buildings. The second experiment is to test the robustness of the BEMR when the testing data is only partially covered by the training data.BEMR correctly identified the best model on 83% of the buildings. The third experiment uses a real building case to validate the proposed framework and the result shows promising applicability and extensibility. The experimental results show that BEMRis capable of adapting to a wide variety of

building types ranging from a service restaurant to a large office, andgives excellent performance in terms of both modeling accuracy and computational efficiency.

Keywords

Building energy consumption; time series forecasting; recommendation system; machine learning;meta-learning;feature reduction;

1 INTRODUCTION

According tothe U.S. Energy Information Administration (EIA),buildings consume nearly half (48%) of the total energy and produce almost 45% of CO_2 emissions in the United States[1]. This drives the need to develop high-fidelity and computationally efficient energy forecasting models for building systemsto ensureoptimal automatic operation, reduce energy consumption, and improve the building's resilience capability to power griddisturbances[2]. Existing building energy models arein general categorized as: physics-based models, hybrid models and data-driven models(Li and Wen 2014).Physics-based modelsemploy thephysical concepts and knowledge of the low level devices and aggregate the mathematical expressions to model the building system. It heavily relies on domain expertise and often is computationally prohibitive[4].Hybrid models use simplified physical descriptions combined with parameter identification algorithms to predict energy consumption. Nevertheless, without a description of the building geometry and materials, it is difficult to estimate the model parameters.In contrast, the emerging technology advancements in the energy industrymake it possible to collect massive amounts of data from sensors and meters, which enable data-driven modeling to unfold the underlying knowledge[5]. As most industrial, institutional, and commercial buildings built after 2000 include a building automation systems (BAS), there is a growing interest to mine valuable information and derive additional insights from data collected.Thedata-driven approach motivates and drives the building energy research in various aspects including estimation of energy consumption[6]–[8], real-time performance validation and energy usage analysis[9], and energy saving operational control[3], [10], [11].A significant advantage of the data driven approach lies in that it considerably reduces the designcycle iteration time for building design and operations, which includes not only simulation, but also analysis of results and optimization of actions based on these results. It allows for fast realizations of the design and operation tasks for any building scenario in an industrial context.Based on the updating cycle and horizon, the load forecast models can also be categorized into short term load forecasting (STLF), medium term load forecasting (MTLF), and long term load forecasting (LTLF)[12]. STLF focuses on the load forecasting on daily basis and/or weekly basis, and MTLF and LTLF are based on monthly and yearly collected data for transmission and distribution (T&D) planning [13], and financial planning, which assist with medium to long term energy management, decision making on the utilities project and revenue management.STLF is important for real-timeenergy operations and maintenance. For daily operations, system operators can make switching and operational decisions, and schedule maintenance based on the patterns obtained during the load forecasting process[14]. To better assist the operations and control strategies development, this study develops a novel STLF methodology for buildings, which provides accurate load forecasts for daily and weekly based energy system management.The model, however, could be viably transformed into MTLF or LTLF, by adding features of economy and land use, and extrapolating the model to longer horizons.

Variousdata-driven methods have been studied and implemented for building load forecasting including 1) statistical methods such as autoregressive, moving average, exponential smoothing [15], state space [16], [17], polynomial regression [18], and 2) machine learning methods such as neural networks[19] and support vector regression [8], [20].Statistical regression models simply build the correlation between the energy consumption and the simplified influential features such as weather parameters. These empirical models are developed from historical performance data to train the models.

Machine learning models are good at building non-linear models and are especially effective on complex applications.

A regression-based approach was testedon the peak and hourly load forecasts of the next 24 hours using Pacific Gas and ElectricCompany's (PG&E) data [21]. The regressionmodel was thoroughly tested and concluded to be superior to the existing system load forecasting algorithmsused at PG&E.In another study, five methods(autoregressive integrated movingaverage (ARIMA) modeling; periodic AR modeling, an extension for doubleseasonality of Holt-Winters exponential smoothing; an alternative exponentialsmoothing formulation; and a principle component analysis (PCA) based method) were compared on 10 load series from 10 European countrieson an hourly interval and 24-hour horizon[22]. They concluded that the double seasonal Holt-Winters exponential smoothing methodoutperformed the others. Another interesting study by Ahmed, Atiya, Gayar, & El-Shishiny (2010) explored machine learning methods. Eight machinelearning models for time series forecastingonthe monthly M3 time series competition data (around a thousand time series) were investigated. These eight are multilayer perceptron, Bayesian neural networks, radial basis functions, generalized regressionneural networks, K-nearest neighbor regression,CART regression trees, support vector regression, and Gaussian processes. They concluded that the besttwo methods turned out to be the multilayer perceptron and the Gaussian process regression.Chirarattananon and Taveekun (2004) developed a model for building energy consumption forecasting based on overall thermal transfer value and concluded that the model does not present good generalizability on some types of buildings, especially on hotels and hospitals. Yik, Burnett, and Prescott (2001) predicted the energy consumption for a group of different types of buildings using a number of physical parameters such as air conditioning system type, year the building was built and geographical information. The resulting model showed high correlation to the detailed simulation model. One novel data-characteristic-driven modeling methodology for nuclear energy consumption was proposed in [26], in which two steps, data analysis and forecasting modeling, were involvedinformulating an appropriate forecasting model in terms of the sample data's own data characteristics. Experimental results showed that "data-characteristic-driven modeling" significantly improves prediction performance compared to all other benchmark models without consideration of data characteristics. However, only timeseries data characteristicsandunivariate forecastingmodels wereexplored in this study. One observation from these extensive studies is model performance varies and is highly dependent on the characteristics of the building systems, which leads the researchers come to inconsistent conclusions regarding the performance of various forecasting models. This concurs with what was found by [27]: he thoroughly reviewed twenty-five years of research and concluded that no algorithm is best for all load forecasting tasks. He suggested that the identification of which methods should be chosen with respect to the situationsshould be done viaexperimental studies.

Notingthat a building system is stochastic, nonlinear and complex[28], research so far has mainly focused on anapproachof trial-and-error or one-size-fits-all. In the cases where little prior knowledge of the building systems is available, previous studies either develop multiple models and identify the outperformer among them, which is computationally expensive and impractical for real-time building energy management and operations,orsubjectively presume one model fits any type of building, suffering from high-bias modeling. In shortterm building load forecasting, the main goal is to minimize the forecasting error with computationally-efficient solutions. Building management control tasks can range from real-time load forecasting and user behavior analysis to predictive building control.For these tasks, the meter data are usually generated at a rate ranging from per minute to per hour. Due to the dynamics of building energy systems and for real-time supervisory purposes, the control and operations should be updated dynamically by analyzing the time series data. This impedes the trial-and-error modeling approach in that the computational complexity for constructing multiple models is unaffordable, especially in the case where data volume is large. In a broader scope, a reduction of the forecasting error ensures the power systems stabilize in balance and assists power market design, operation, and security of supply [29]. These drive the need for a general framework for shortterm building load forecasting, which satisfies both the time constraint driven by real-time building operations and control, and the fidelity constraint which calls for high-accuracy load forecasting. The general building load forecasting framework would be beneficial in dealing with heterogeneous building load forecasting tasks for most commercial utilities and market participants.Taking into account the above, we develop a Building Energy Model Recommendation (BEMR) system for shortterm load forecastingmotivated by the meta-learning concept.Meta-learning has gained increasing attention and has been successfully applied in diverse research fields including gene expression classification [30], failure prediction [31],gold market forecasting(Zhou, Lai, & Yen, 2012), and electric load forecasting [33], just to name a few. Meta-learning is a machine learningalgorithmthat explores the learning process and understands the mechanism of the process, which can be re-used for future learning. The objective is to build a self-adaptive automatic learning mechanism that connects the meta-data (e.g., the characteristics of the problems) with the model performance. As a result, the best performing model can be identified via the meta-data directly and thus significantly saving the model training process.

Earlier efforts on meta-learning for forecasting mainly focused on rule-based approaches. For example, [34]weighted four candidate models using 99 derived rulesfrom human experts' analysis. The weight of each model is modified based on the features of the time series. One potential issue of this approach is the knowledge acquired from human experts may not be easily accessible. Prudêncio & Ludermir (2004) used a decision tree on a stationary time series with two candidate algorithms,exponential smoothing witha neural network,and NOEMON, on the M3-competition time series, for ranking three candidate models: random walk, Holt's smoothing, and auto-regressive. They concluded both casestudies revealed satisfactory results, taking into account the quality in the selection andthe forecasting performance of the selected models. Wang, Smith-Miles, & Hyndman (2009) generated a decision tree on the induced rules from univariate time series data characteristics, where four algorithms: Random walk, smoothing, ARIMA, and neural network, were selected as candidates. They were able to draw recommendations and suggestions on the conceptive, categorical and quantitative rules.The meta-learning system based on a large pool of meta-features proposed by [37] was shown to outperform many approaches of the NN3 and NN5 competition entries.Marin Matijaš, Suykens, & Krajcar (2013) proposed a meta-learning system for load forecasting based on multivariate time series, in which 65 load forecasting tasks in Europe were tested and lower forecasting errors were observed compared to 10 well-known forecasting algorithms.

Note that the literature reviewed above all attempt to gain knowledge from time series data to generate rules which define the relationship between the meta-features and the model performance. While promising for the problems examined, building systems are inherently nonlinear, diverse and complex due to the heterogeneity among multiple interconnected factors, e.g., internal factors, social factors and weather factors [28]. For buildings, especially large and complex ones, simplifications of model formulations and lack of physical knowledge may lead to poor forecast accuracy. Therefore, the meta-knowledge characterization should not solely be collected from the building's operational data, such as energy consumption univariate time series, but also the building's physical features.

We conclude that a generalized intelligent system for building energy model recommendation, which incorporates both building data-characteristic and physical-characteristic meta-features is currently lacking and this research attempts to fill this gap motivated by the research success from [39].Specifically, we employ a two-stage meta-learning approach for BEMR. It first trains multiple models on the existing buildings to obtain the model performance. Next, the features and/or meta-features are derived from the existing building instances in association with the respective performancesfor making recommendations on the new building. The BEMR framework developed in this study can be used on development and selection of models for building energy modeling and forecasting, as well as building optimal operation and real-time control.

In developing BEMR,the firstnotable challenge is that building data is of high dimension in both the temporal and spatial domains. Building energy consumption is influenced by many factors: internal factors such as building structure and physical characteristics, the sub-system components like equipment schedule and operations on HVAC systems, occupants and their behavior, and external factors such as natural environments, weather conditions, and economies. Therefore, meta-features are introduced to depict the operational data, and the physical features of the buildings are gathered as additional descriptive knowledge. We hypothesize the inclusion of the heterogeneous features should increase the generalization of BEMR for diverse buildings in different operating conditions. Next, six statistical and machine learning data-driven models are explored and included in BEMR: Kriging, support vector regression (SVR),radial basis function (RBF), multivariate adaptive regression splines (MARS), artificial neural network (ANN) and polynomial regression (PR). These models are chosen due to their extensive use in surrogate modeling applications [40] and their good theoretical and experimental performance on energy system applications [41], [42].The third effort in BEMR is to collect the building instances as the training sources. Considering that both the building type (internal factors) and climates (external factors) have effects on energy consumption profiles,48 (8 building types on 6 climate zones) simulated commercial and residential reference buildings developed by the Department of Energy (DOE)are collected.Last, ANN is chosen as the meta-learner to develop the associations between the meta-features derived from the building instances and the model performance so the best model is identified. Three sets of experiments are conducted using leave-one-out cross validation. The first experiment is to test the performance of BEMR on regular shortterm daily and weekly forecasting. Experiment results show that among the 48 buildings, BEMR is able to identify the best model for 43 buildings (accuracy: 90%) and the difference of the mean ofthe normalized root mean square error (NRMSE)from the ground truth is within 2%. The second experiment is to validate the robustnessof BEMRwhenthe test data is only partially covered by the training data, and we call it extrapolation validation. Among the 48 buildings, 40 (accuracy: 83%) correct model recommendations are made and the difference of mean NRMSE from the ground truth is within 3%. Moreover, the computational cost of the system is significantly lower than traditional trial-and-error approaches, which decreases forecast time from the order of minutes to seconds.The third experiment is to validate the proposed framework on a real building case, which is located in Ankeny, IA. The result shows that the proposed BEMR is capable of making reliable recommendations for a real building energy forecast.

The paper is constructed as follows: Section 2introduces the proposed methodology; Experiments and results are discussed in Section3; finally, a discussion of the conclusion and future work is given in Section 4. The appendix gives a brief discussion on the data-driven forecasting algorithms.

2 BUILDING ENERGY MODEL RECOMMENDATIONSYSTEM

In this research, we propose a Building Energy Model Recommendation System (BEMR) for short-term building energy consumption forecasting. BEMR is a two stage framework. As shown inFigure 1,the first stage is to establish the instance repository to connect the learning instances with a forecasting models'performance; next, both building physical features and operational meta-features are derived and connected with the model performances so the model recommendation can be made.

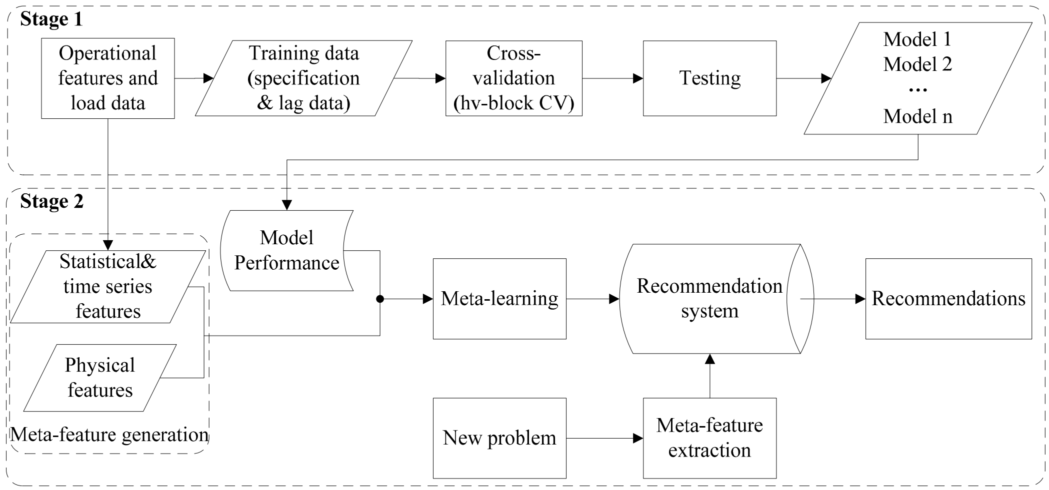


Figure 1Framework of Building Energy Model Recommendation (BEMR)System

2.1 Stage I: Building Learning Instance Repository

Eight types of commercial and residentialbuildings are selected fromthe DOE simulated reference buildings whichare identified as the most prevalent building types [43] in the United States. Considering the significant impact of climate on the energy consumption profile, each building type is simulated ateach of sixselected locations which correspond to the climate zones discussed in ASHRAE 90.1 - 2004[44]. These locations are San Francisco, CA; Boulder, NV; Phoenix, AZ; Houston, TX; Miami, FL; and Baltimore, MD. As a result, the building repository includes a total of 48 simulated buildings (8 types, in 6 locations). The corresponding TMY3 (typical meteorological year) weather data sets [45] are adopted as the weather data source for the simulation models.

2.1.1 Training Data Selection

The STLF process heavily relies on the weather information and ambient environment.When the parameters are estimated, the weather information is extrapolated to forecast the load. Much research[4], [20]has looked at the most suitable features for load forecast problems. They try to explain the causality of the electric load consumption. In STLF, the electric load is generally driven by nature and human activities. Nature is usually represented by weather variables, e.g., temperature and humidity, while the human activities are usually represented by the calendar variables, e.g., occupancy and business hours. High-dimensional feature spaces result in unnecessary complication in building forecasting models and thus impede the optimization process.To alleviate this concern, our features are selected based on the work ofEisenhower et al.(2012), in which the sensitivity analyses were conducted to identify the most influential features for the energy output generated from the EnergyPlus simulation models.They were adopted todevelop the meta-model and the followingoptimization model for energy management operations. Seventop influential variables, which are all temperature and human activity related, were selected to build areduced form of meta-models. On the foundation of their work,12 operational features are initially selected from over 600 features in the simulation models, including (1) outdoor air dry bulb temperature; (2) outdoor air relative humidity; (3) outdoor air flow rate; (4) diffuse solar radiation rate; (5) direct solar radiation rate; (6) zone people occupant count; (7) zone air temperature; (8) zone air relative humidity; (9) zone thermostat cooling set point temperature; (10) building equipment schedule; (11) building light schedule; (12) HVAC operation schedule. In addition, since periodicity is one main characteristic in electricity load time series, two categorical variables, Day and Time are added to the study. Given these 14 features, we then conduct principal component analysis (PCA)[46] to explore the multicollinearity among the features for robust forecasting model development. It is observed that feature 11 (building light schedule) and feature 12 (HVAC operation schedule) are highly correlated with feature 9 (zone thermostat cooling set point temperature). Therefore, these two highly collinear variables are removed from the study. We further assess the correlation between each remaining feature and the response variable using Pearson's correlation coefficient. It is observed that all the features are significantly correlated to the response variable (the absolute correlations are all above the threshold correlation, 0.195, to reject the null hypothesis that the two variables are not correlated). Note categorical variables are excluded in the multicollinearity test and the correlation test. Finally, tenbuilding operational features andtwocategorical variables are selected (Table 1).

Table 1.TenSelectedBuilding Operational Features and twoCategorical Variables

	Building Variables	**Variable Type [range]**
1	Outdoor Air Drybulb Temperature (℃)	Continuous
2	Outdoor Air Relative Humidity	Continuous on [0,1]
3	Outdoor Air Flow Rate	Continuous
4	Diffuse Solar Radiation Rate (W/m^2)	Continuous
5	Direct Solar Radiation Rate (W/m^2)	Continuous
6	Zone People Occupant Count	Integer
7	Zone Air Temperature (℃)	Continuous
8	Zone Air Relative Humidity	Continuous on [0,1]
9	Zone Thermostat Cooling SetPoint Temperature (℃)	Continuous
10	Building Equipment Schedule Value	Continuous on [0,1]
11	Day of Week	Integer on [1,7]
12	Time of Day	Integer on [1,48]

Besides the features discussed above, all the buildings (simulation models) apply typical equipment control strategies for chillers and fans. In fact, no matter how the subsystems/devices are controlled, their operations will be reflected in the training data. Our models should be able to capture these operation characteristics in the model training process.The objective of this study is to provide whole building level STLF models for building operation and control. As a result, only the building level features are selected. The detailed sub-system level and device level operation are not studied in this paper.

For the features, both specification data and lagged data are collected in the training data set. Specifically, let c be the periodicity of the seasonality, n be the number of lags, and$t$ be the current time data index, then the specification data indices are t, $t - c$, $t - 2c$, while the lagged data indices are t-1, t - 2,…, t–n. For example, assume the current time t is 12 pm on a day, possible lagged data indices are 11:30 pm, 11 pm, 10:30 pm, etc. (given data are collected every 30 minutes), and possible specification data indices are 12 pm in the past few days (c=24 hrs.). This is motivated by the "Similar Days technique" in [47]that a particular load on the same day of the weekshouldbehave similarly, given similar weather and other conditions. Several researchers have pointed out the superior performance of specification models over traditional models which arebuilt solely on lagged data (Crespo Cuaresma, Hlouskova, Kossmeier, & Obersteiner, 2004).

2.1.2 Cross validation

It is worth noting that in traditionalforecasting, a common practice is to reserve some data toward the end of each time series fortesting, and to use earlier time series data for training. One potential issue is that the data are not fullymade use of due to a lack of cross-validation, and the resultingmodel may suffer from over-fitting. Meanwhile, for time series data it may not be appropriate to directly apply traditional cross-validation,which randomly splits the data into training and testing datasets.Theoretical problems with respect to temporal evolutionary effects and data dependenciesare encounteredwhenthe fundamental assumptions ofcross-validation might be invalidated.Racine (2000) proposes "*hv*-block" cross-validation which is asymptotically optimal. It is consistent for temporally dependent observations in the sense that the probability of selecting the model with thebest predictive ability converges to 1 as the total number of observations approaches infinity. The basic idea is to place restrictions on therelationship between thetraining set, validation set, the size of an*h*-block, and the sample size. We canthereby obtain a consistent cross-validating model selection procedure for the process.

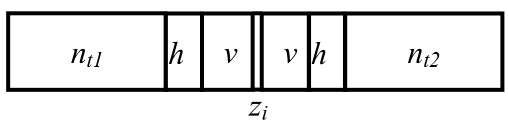


Figure 2"*hv*-block"Cross-validation Illustration

As shown inFigure 2, given an observation z_i, wefirst remove v observationson either side of itto obtain a validation set of size$2v+1$.We thenremove another h observations on either side of thisvalidation set with the remaining $n-2v-2h-1$ observations forming thetraining set. Thevalue of v controls the size of the validation set with$n_v= 2v+1$. *The* value of h controls the dependence of the training set of size$n_t= n-2h-n_v$ and the validation set of size n_v. For guidance on appropriate selection on h and v, please refer to [48] for details.

For illustration, Figure 3showsthe design for cross-validation on a single day test. Take Friday as an example, and let's define it as F_0, and the unit of lag being a day, with n being 6 days, and c being 7 days. Therefore, the training data consists of six days of lagged data (Thursday, Wednesday, Tuesday, Monday, and Sunday on the same week of test data, and Saturday from the previous week) and three days of specification data (three Fridays from the last three weeks, F_1, F_2, F_3). Based on the"*hv*-block" cross-validation approach, the training data are cross split into 4 training and validation folds. In each fold, the size of validation data n_v and the block h are set as one day, and the rest of data is kept aside as training data.

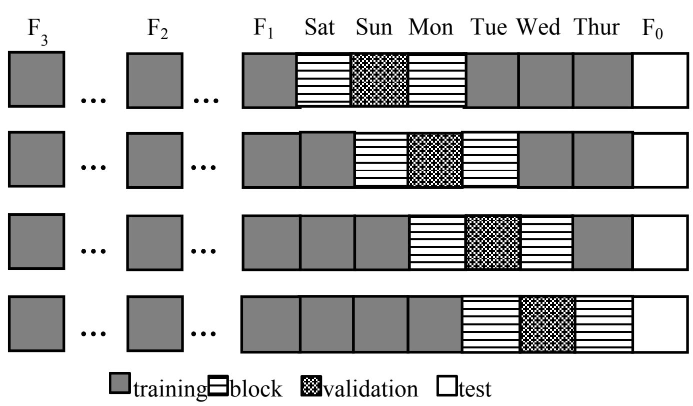


Figure 3Cross-validation of Training Data Split

2.1.3 Forecasting Model Performance Evaluation

In BEMR, six data-driven models are explored including Kriging, support vector regression (SVR),radial basis function (RBF), multivariate adaptive regression splines (MARS), artificial neural network (ANN) and polynomial regression (PR). To make the recommendation, the first step is to evaluate and validate the model performance using available building energy data. The performance is measured using Normalized Root Mean Square Error (NRMSE), where

$$NRMSE = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N}} / (y_{max} - y_{min}), \tag{1}$$

and y is the true value of the building energy consumption and $\hat{y}$ is the forecast value.

In summary, stage I of the BEMR is providing the base repository which consists of 288 models (8 building types, 6 locations, 6 data driven models) and the respective forecasting performance (measured by NRMSE). This enables the implementation of the meta-learning strategy which is discussed in the next section.

2.2 Stage II: Meta-level Learning

2.2.1 Meta-Feature Extraction

Meta-features, which characterize the entire dataset for meta-level induction learning, arean abstraction of knowledge extracted from the dataset.Three types of meta-features are devised, including physical features, statistical features and time series features. Table 2 summarizes the seven physical features of the buildings.

Table 2BuildingPhysical Features

Feature #	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Building Type	**# of stories**	**Area(m^2)**	**Roof Type**	**Wall Type**	**Window Type**	**Cooling**	**Space Type**
Large Office	12[1]	46,320	IEAD[2]	Mass	Fixed	Chiller, water-cooled	Non-residential
Medium Office	3	4,982	IEAD[2]	Steel Frame	Fixed	Packaged DX[3]	Non-residential
Small Office	1	511	Attic Roof	Mass	Fixed	Packaged DX[3]	Non-residential
Supermarket	1	4,181	IEAD[2]	Mass	Fixed	Packaged DX[3]	Non-residential
Full Service Restaurant	1	511	Attic Roof	Steel Frame	Fixed	Packaged DX[3]	Non-residential
Hospital	5[1]	22,422	IEAD[2]	Mass	Fixed	Chiller, water-cooled	Residential for patient rooms
Large Hotel	6[1]	11,345	IEAD[2]	Mass	Operable in guest rooms	Chiller, air-cooled	Residential for guest rooms
Midrise Apartment	4	3,135	IEAD[2]	Steel Frame	Operable	Packaged DX[3]	Residential

[1] Plus Basement.
[2]Built-up flat roof withinsulation entirely above the roof deck.
[3] Packaged Direct-expansion (DX) equipment.

Other than the sevenphysical meta-features, nine statistical meta-features similar to (Matijaš, 2013; Lemke & Gabrys, 2010)are derived from the operational featuresfromTable 1 and the energy consumption data:

(S1) Min: e.g.,the minimum of load over a time period

(S2) Max: e.g., the maximum of load over a time period
(S3) Mean: e.g., arithmetic average of load over a time period
(S4) SD: e.g., the standard deviation of load over a time period
(S5) Skewness: evaluates the lack of symmetry, taking the load as an example, Y_i is the load of timeperiod i, and $\overline{Y}$ is the mean of the load over a period of time, skewness is derived as:

$$E\{[(\boldsymbol{Y_i} - \overline{\boldsymbol{Y}})/\boldsymbol{Std.}(\boldsymbol{Y_i})]^{\boldsymbol{3}}\}, \boldsymbol{i} = \boldsymbol{1}, \ldots, \boldsymbol{N}, \tag{2}$$

(S6) Kurtosis: evaluates the flatness relative to a normal distribution. Again, taking the load as an example

$$E[(\boldsymbol{Y_i} - \overline{\boldsymbol{Y}})^{\boldsymbol{4}}]/(\boldsymbol{E}[(\boldsymbol{Y_i} - \overline{\boldsymbol{Y}})^{\boldsymbol{2}}])^{\boldsymbol{2}}, \boldsymbol{i} = \boldsymbol{1}, \ldots, \boldsymbol{N}, \tag{3}$$

(S7) Q1: e.g., 25% quartile of load, which is the lower quartile of load.
(S8) Q2: e.g., 50% quartile of load, which is the median of load.
(S9) Q3: e.g., 75% quartile of load, which is the upper quartile of load.

In addition, considering the building system is dynamic and non-linear, we introduce four time series meta-features to describe the temporal characteristics of the building energy data.

(T1) Ratio of local extrema: Ratio of local minima and maxima within a given neighborhood, taking the load as an example, it measures the percentage of load oscillation.
(T2) Non-linearity: A number of surrogate data is generated from the null hypothesis that the series is linear, and the derived estimate of the original time series data is compared to the ones generated from the surrogate data to check the non-linearity [49].
(T3) Cut-off lag of ACF: The autocorrelation function (ACF) is the collection of the autocorrelation coefficients, which indicate the covariance between observations with any lag. In this study, a lag of 30 autocorrelation coefficients is calculated.
(T4) Cut-off lag of PACF: Similarly, a lag 30 of the partial autocorrelation function (PACF)is used to derive the coefficients.

As a result, we derive nine statistical meta-features for each of theten building operational data and the energy consumption data (99 meta-features in total). Additionally, four time series meta-features on the energy consumption data are derived. With the seven building physical features a total of 110 features (meta-features) are used for meta-learning.

2.2.2 Meta-learner

[50], [51] indicate that a powerful artificial intelligence-based model is more preferable than traditional statistical models. Therefore, we use an ANNas the meta-learner, considering correlation between the meta-features and nonlinear patterns brought by the complexity and heterogeneities of different building scenarios (noises within meta-features)might impair the modeling power of the learner. The parameter settings of the meta-learner ANN areas follows: the hidden layer size is tuned within the range of [10, 20], and the transfer functions are selected between radial basis and log sigmoid. Note that the proposed meta-featuresare tentatively selected in hoping that they could effectively represent the dataset. However, the number of features is more than twice the number of problems, which may impair the predictive power of the meta-learner. This is known as the "*Hughes effects*"[52]. As a result, we propose to use an advanced feature reduction technique to address the curse of dimensionality. Specifically, singular value decomposition (SVD) is of interest in this research due to its known performance on noisefiltering and dimensionality reduction.It is a factorization of a real matrix $X \in R^{m \times n}$, $m \geq n$,

$$X = USV^t, \tag{4}$$

where $U \in R^{m\times m}$ and $V \in R^{n\times n}$ are orthogonal matrices and $S \in R^{m\times n}$ is a diagonal matrix. A rank-k ($k \ll min\,(m,n)$) matrix C is defined as the best low-rank approximation of matrix X if it minimizes the Frobenius norm of the matrix $(X - C)$, which is known as the Eckart–Young theorem [53]. This approximation matrix can be computed by SVD factorization and keeping the first k columns of$U$, truncating S to the first k diagonal components, and keeping the first k rows ofV^t. This results in noise reduction by assuming the matrix $X$is low rank, which is not generated at random but has an underlying structure.

2.2.3 BEMR Performance Evaluation

Given the predicted rankings of the six models' performance from the recommendation system, two evaluation metrics are introduced to evaluate the meta-learning performance: The Spearman's rank correlation coefficient (SRCC) and success rate.

The Spearman's rank correlation coefficient [54] is employed to measure the agreement between recommended rankings and ideal rankings on a forecasting problem. For two samples of size *N*, the rank coefficient is computed as

$$\rho = 1 - 6 \cdot \frac{\sum_{i=1}^{N} d_i^2}{N(N^2-1)}, \quad (5)$$

where $d_i = r_i - l_i$, andr_i and $l_i$are the recommended rankand the ideal rank on the i^{th} sample. In this case, the sample size *N* is the number of candidate forecasting models. Thevalue of 1 represents perfect agreement while −1, perfect disagreement. A correlation of 0means that the rankings are not related, which would be the expected score of the randomranking method.

The percentage of exact matches between ideal best performer and recommended best performer over all problems is defined as Success Rate. This is to evaluate the "precision" of the meta-learning performance. As a matter of fact, in the case of forecasting, users are sometimes more concerned if the recommended best performer (top 1) matches the ideal one, so only one model is built and computational efficiency is ensured. Therefore, besides the Spearman's rank correlation coefficient, the success rateis alsoproposed to comprehensively evaluate the performance of the meta-learning system.

3 EXPERIMENTS AND RESULTS

In this study, we investigate the cooling electricity consumption of buildings in the summer time. Simulation data are obtained by simulating the reference building energy consumption models for one month inJuly. The data are generated athalf-hour granularity using DOE's EnergyPlus[55] simulation software, which yields 48 data points on each day, 1,488 data points for a month.Three forecasting cases are tested respectively: (1) Single day and a one week test, (2) an extrapolation test, and (3) a real building validation test.

3.1 Experiment I

In this set of experiments, we test the performance of the proposed BEMR to forecast the building cooling load for each day of the last week and the whole last week of July, respectively.The single day test and one week test correspond to short-term load forecasting on a daily basis and a weekly basis. In the one week test, since the training data is scarce compared to the size of test case, we apply a traditional validation technique, where the first 80% of the data is used fortraining and the last 20% of the data is used for validation.

Figure 4 displaysabar chartof the mean ofthe NRMSE measures of the best forecasting model across the 48 problems oneach test case.Except for the test on Sunday, the means of the best NRMSE are evenly distributed from 0.020 to 0.035, while the best performance on Sunday is significantly worse than those on other days.To explain this observation, we may refer to the time series plot of the energy consumption inJuly. See Figure 5(a) for the weekly energy time series plot of thelarge office building in San Francisco, CA, which shows that the cooling load of Sunday is significantly less than other weekdays. The sudden decline may be due to the fact thatmost people don't come to work on weekends thus less cooling load is required. On the other hand, due to its significantlydifferent pattern from the weekdays, data available for forecastingthe energy consumptions for Sunday is scarce.This implies more training data with similar patterns are needed for energy forecasting on weekends. Figure 5(b) shows the time series plot of the cooling load of the same type of building located in Phoenix, AZ. Compared to plot (a),similar daily and weekly quasi-periodic behaviors are observed on the energy consumptions, with approximately constant variance and repeated patterns. However,the cooling load of the large office in Phoenix is on average one-tenth more than that in San Francisco, which is to be expected due to the hot summer in Phoenix. Figure 5(c),which displays the cooling load time series plot of a full service restaurant in Phoenix, AZ, shows a markedlydifferent behavior. The daily cooling load presents a stable pattern while the weekly periodicity is not as significant. This is likelydue to the fact thatrestaurants are usually open seven days a week.Moreover, it is observed that the magnitude of the energy consumption in a restaurant is significant lower than that in a large office. These validate our proposition that cooling energy consumption is impacted by combined social factors, weather conditions and building types.

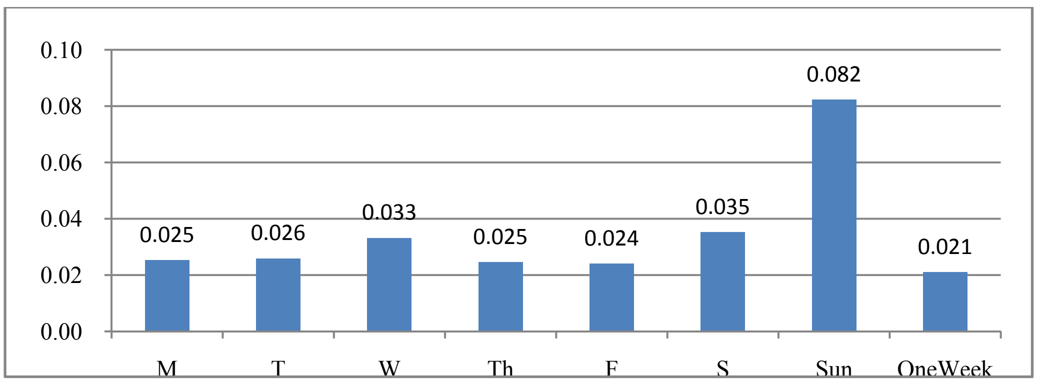
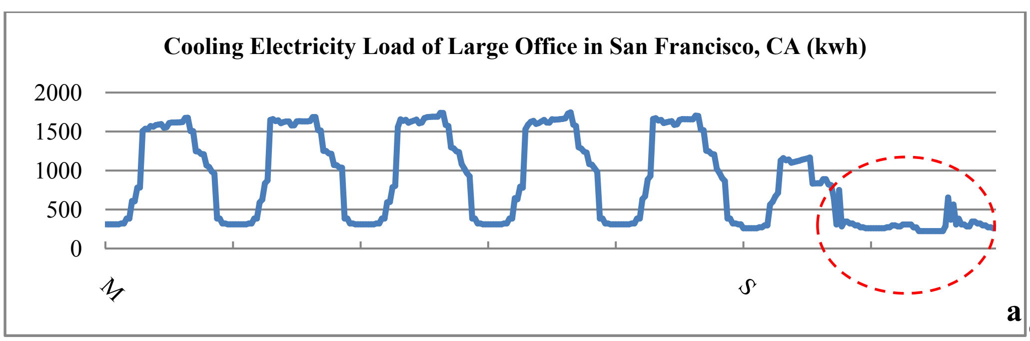


Figure 4Test Case I: Bar Chart ofMean of Best NRMSE across 48 Problems on Each Test Case



c

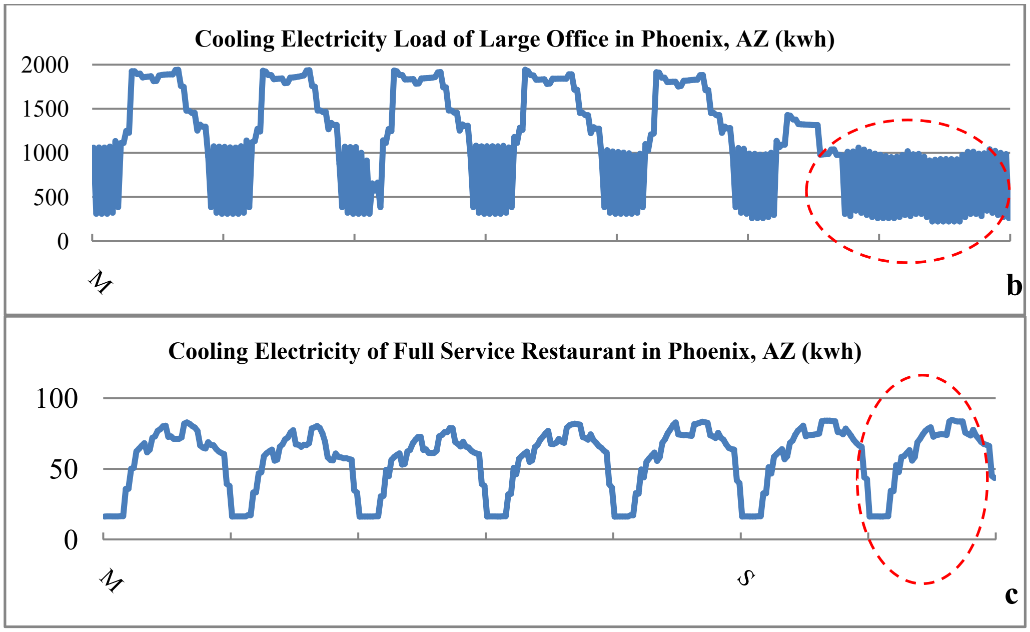


Figure 5Weekly Cooling Electricity Load (Kwh) Time Series Plot of (a)Large Office in San Francisco, CA; (b)Large Office in Phoenix, AZ; (c) Full Service Restaurant in Phoenix, AZ.

Figure 6 andFigure 7 present the meta-learning performance in terms of success rate and SRCC. Table 3summarizes the statistics of the above two performance measures. The average success rate amounts to 90%, which means almost 43 out of 48 problems are correctly assigned with the best model. Again, Sunday has the lowest success rate due to its different patterns from other days. In another words, its meta-features are less similar to others'causingdifficulty inmeta-learning.It is also observed that all the performance measures on the one week test are slightly better than those on the single day, however, notice that the training cost for the one week forecast is much higher than the single day forecast due to the higher training size. Be advised that there arealways trade-offs between the computational cost and model performance, which is worth consideration when selectingthetraining and testing sizes.Please refer to Racine (2000) for a discussion on training, validation and testing sample size selection for time series forecasting using "*hv*-block" cross validation. In addition, the mean SRCC is around 96%, which implies high agreement between the predicted rankings of the recommendation system and the true rankings of the six forecasting models.

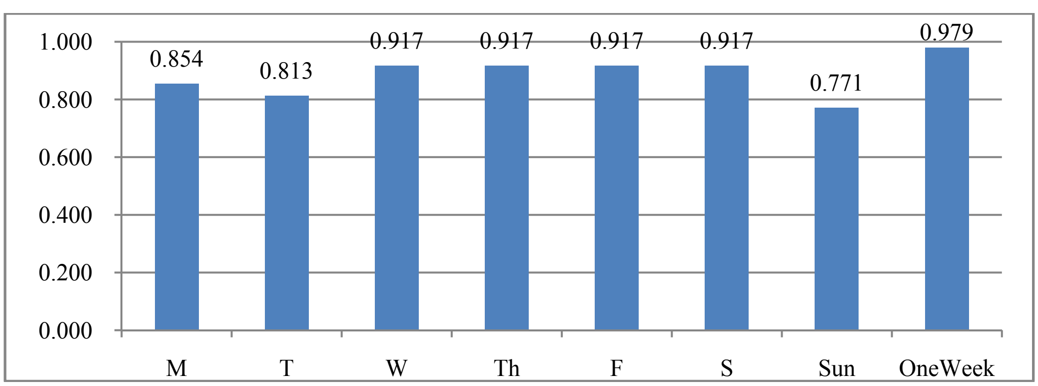


Figure 6Test Case I: Bar Chart ofMeta-learning Success Rate

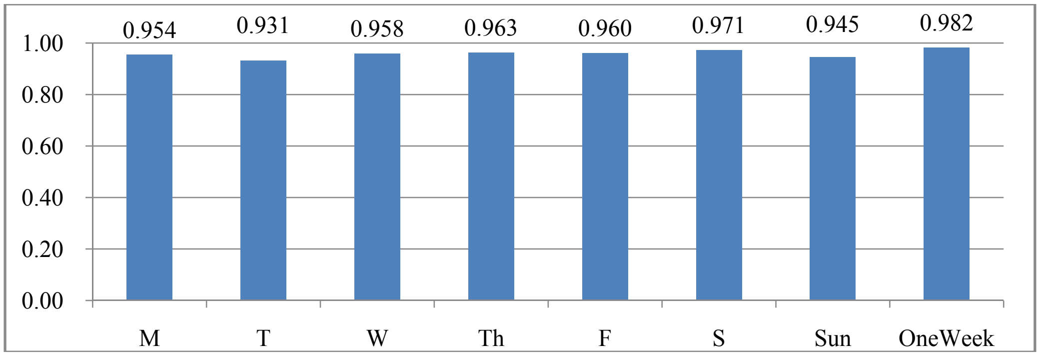


Figure 7Test Case I: Bar Chart ofMeta-learning SRCC

Table 3Test Case I: Statistics on Meta-learning SRCC, Success Rate and # of Successesacross 48 Problems

Test Date	M	T	W	Th	F	S	Sun	OneWeek	Mean
Spearman Correlation Coefficient	0.954	0.931	0.958	0.963	0.960	0.971	0.945	0.982	0.958
Success Rate	0.854	0.813	0.917	0.917	0.917	0.917	0.771	0.979	0.900
# of Successes (out of 48)	41	39	44	44	44	44	37	47	43

3.2 Experiment II

In this set of experiments, we test the extrapolation capability of the proposed BEMR. We sampled four days: Monday, Wednesday, Friday and Sunday of the last week, toforecast the building cooling load, while the training data is the building cooling load of the first week. Notice that by observing the energy data, some of the features of the last week are out of the range covered by the training data of the first week. For example, the average range of the difference between the maximum and minimum outdoor temperature among all the buildings in the first week is around [24, 35] ℃, while it is around [22, 39] ℃ in the last week. The temperature gap in the training data allows us to test the extrapolation capability of the forecasting models and the recommendation system performance under uncertainties.

Figure 8 displays abar chart of the mean of the NRMSE measures of the best forecasting model across 48 problems on the second test case. An attractive finding is that the best forecasting performance on extrapolation is only slightly inferior to regular forecasting. This can be observed by noting that the difference between the mean values in Figure 4andFigure 8 is around 0.01. This indicates the best forecasting model generally is able to give a reliable forecast even though a time gap exists between the forecast horizon and the energy data at hand. Therefore, energy users and utilities can have confidence in the extrapolation predictions to pre-plan and make decisions in advance, which enables energy savings and cost reductions.Figure 9 displays a box plot of the mean of the NRMSE on the single day, one week and extrapolation tests across six forecasting algorithms. It is observed that the variance of the mean NRMSE for the tests on one day tests of Friday, Saturday, Sunday and the Sunday extrapolation are greater than other days, which indicates that the performance of different forecasting models vary significantly to each other on these days. This may be caused by the dates beingweekends, or quasi-weekend (Friday), when the energyusage patterns are different from regular weekdays.

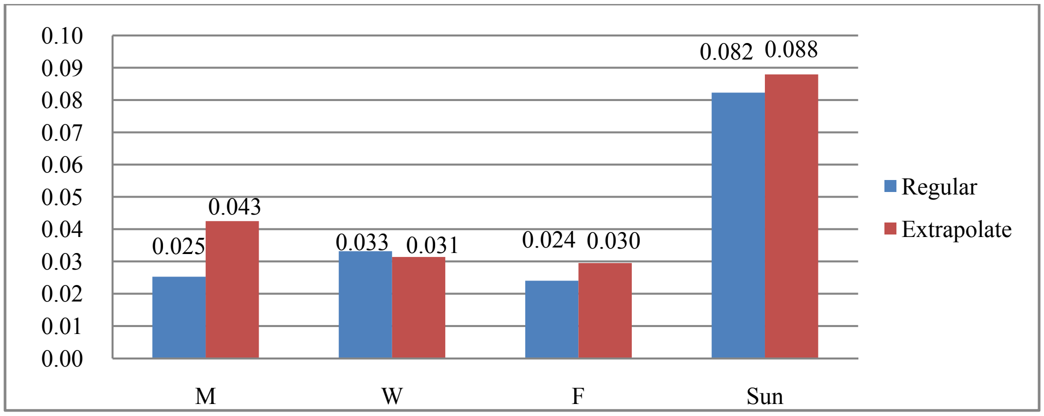


Figure 8Test caseII: Bar Chart of Mean of Best NRMSE across 48 Problems on Each Test Case

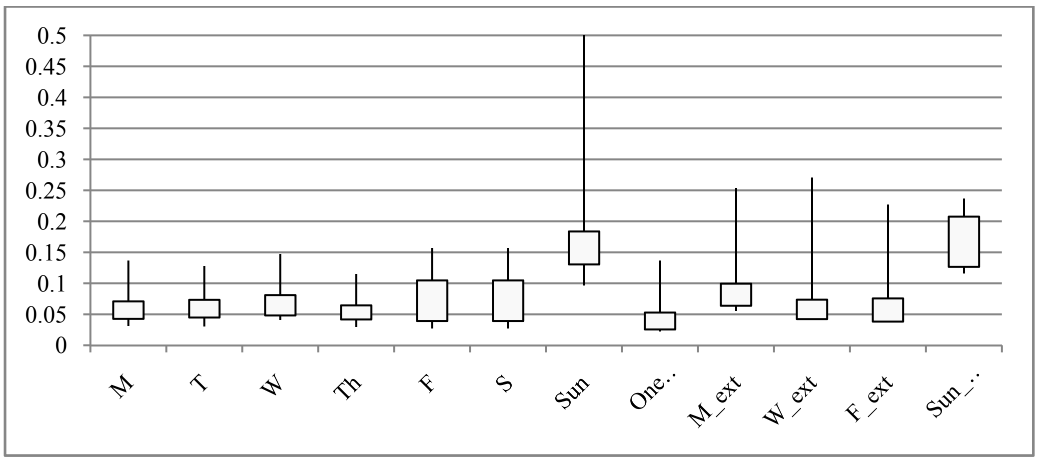


Figure 9 Box Plot of Mean of NRMSE on Test Cases I&II

Figure 10 and Figure 11 present the meta-learning performance in terms of success rate and SRCC on the second test case. Table 4 summarizes the statistics of the above two performance measures. Similar to the comparison result on the best forecasting model performance, all three performance measures are slightly inferior to regular forecasting. The mean SRCC still remains above 94%, and the average successful recommendations are almost 40 out of 48, which is acceptable. Table 5 gives a comparison between the ground truth and the recommendation system of the three test cases based on the mean of the best NRMSE across 48 problems. It is shown that the average discrepancy between the recommended model and the true best model performance is within an error of 0.02, which reveals the proposed system is highly capable of making correct recommendations.

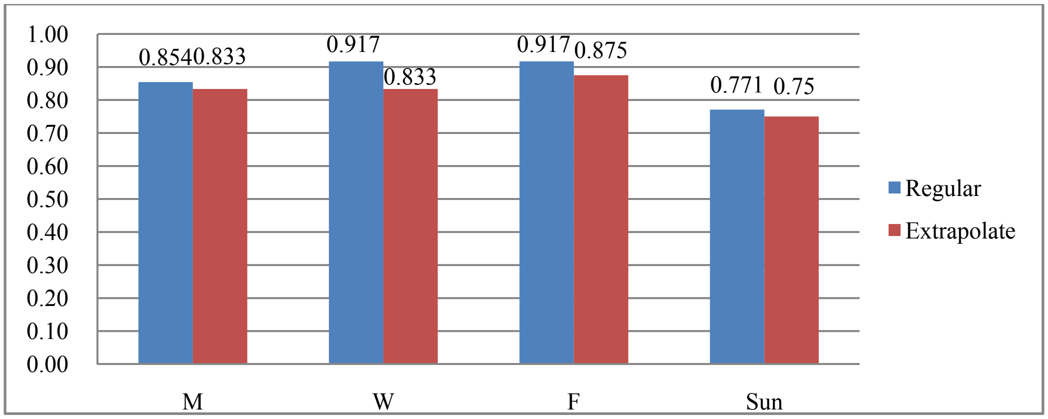


Figure 10Test caseII: Bar Chart ofMeta-learning Success Rate

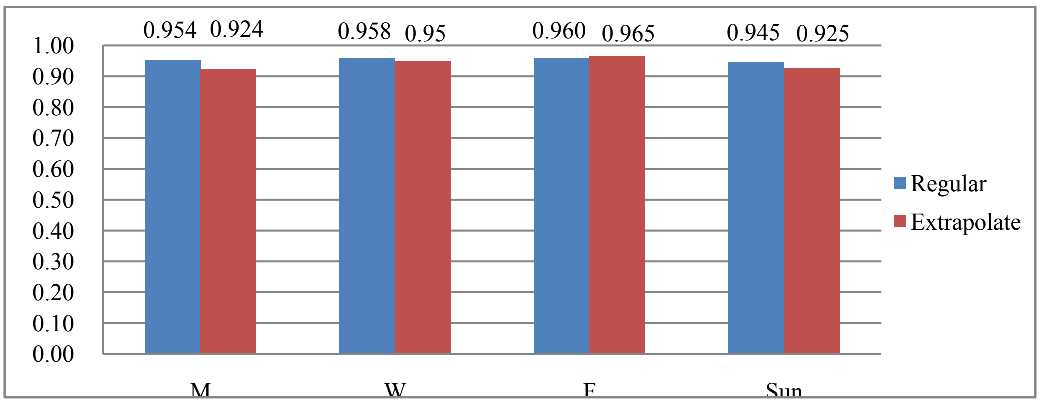


Figure 11Test case II: Bar Chart ofMeta-learning SRCC

Table 4Test caseII:Statistics on Meta-learning SRCC, Success Rate and # of Successes across 48 Problems

Test Date	**M_ext**	**W_ext**	**F_ext**	**Sun_ext**	**Mean**
Spearman Correlation Coefficient	0.924	0.950	0.965	0.925	0.941
Success Rate	0.833	0.833	0.875	0.750	0.833
# of Successes(out of 48)	40	40	42	36	40

Table 5Comparison between Ground Truth and Recommendation System on Mean of Best NRMSE across 48 Problemson Each Test Case

Test Date	**M**	**T**	**W**	**Th**	**F**	**S**
True Best	0.025	0.026	0.033	0.025	0.024	0.035
Recommend	0.026	0.029	0.034	0.026	0.025	0.036
Test Date	**Sun**	**OneWeek**	**M_ext**	**W_ext**	**F_ext**	**Sun_ext**
True Best	0.082	0.021	0.043	0.031	0.030	0.088

Recommend	0.092	0.021	0.045	0.035	0.031	0.092

Table 6summarizes the mean and standard deviation of the computational cost (in seconds) of the six models on an Intel i5 CPU 16G computer. As seen, PR is the most computationally efficient model, followed by RBF and Kriging. The least efficient algorithm is SVR, which takes more than 5 minutes on average to solve each problem. The variance of the computational costs among different models implies that a trial-and-error method is not an efficient approach for solving heterogeneous energy forecasting problems, especially when the number of problems at hand is large and the problems have different levels of complexityand heterogeneities. By summing the solution times of all six models, it is easy to see why a trial-and-error approach for these types of problems is costly. By introducing the automatic model recommendation using a meta-learning approach, the computational cost for forecasting reduces from an order of minutes to seconds.

Table 6Mean and Standard Deviation of the Computational Cost (in seconds) of the Six Models across 48 Problems

Statistics	**Kriging**	**SVR**	**RBF**	**MARS**	**ANN**	**PR**
Mean	2.75	324.94	0.68	202.79	10.44	0.28
Std.	0.27	151.29	0.08	119.22	1.50	0.08

The promising performance indicates that the proposed ANN based meta-learning recommendation system is capable of accurately recommending not only the best model but also the ranking of the models.Thisprovides more freedom for users to select either one or several models, such as building an ensemble of multiple models[56]. Moreover, it can be concluded that the meta-learning approach can achieve both high prediction accuracy and high computational efficiency on heterogeneous forecasting problems.

3.3 Experiment III

In this experiment, we test and validate the proposed BEMR using a real commercial building at the Iowa Energy Center. The building operation data is acquired from ASHRAE 1312 [57].The target is a small size commercial building with an experiment area and common office area. The total floor space of this building is 855.5 m^2. The area of each test room is 24.7 m^2. The percentage of exterior window area to exterior wall area is 54 % for each exterior zone. A built-up roof with insulation is constructed above the roof deck. The zone thermometers are located on the center of the internal wall (shown as the blue box on the floor plan in

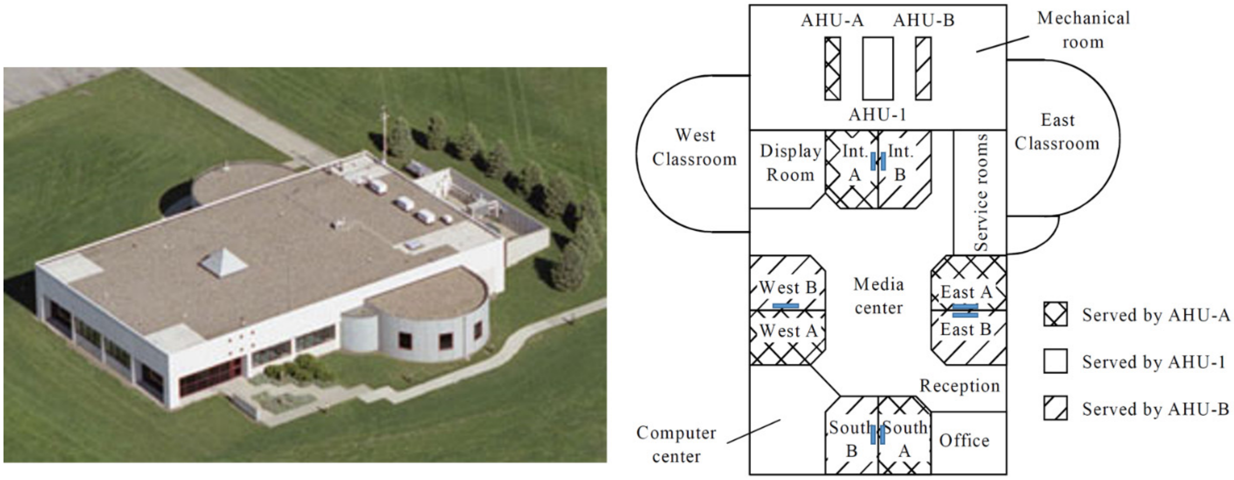


Figure **12**). The location of the sensor is 1.21 meters from the floor. Two Variable Air Volume (VAV) air handling units (AHU) are used for the two experiment systems (A and B) in the experiment area. Both of these AHUs are equipped with dual (supply and return) variable speed fans and are operated similarly to that in a typical commercial building. More details about this building can be found at [58]. In the ASHRAE 1312 experiment, both AHU-A and AHU-B were used. However, AHU-A was used for faulty test and AHU-B was used for regular operation test. As a result, the summer (August and September) test data from AHU-B (system B) was used in this study. Similar to the subsystem operation schemes in experiment I and II, the chilled water temperature set point was 7.2 $^{\circ}$C, the supply air temperature set point was 12.7 $^{\circ}$C, the supply air pressure set point was 9.6 kPa, and the zone temperature heating and cooling set points at occupied hours (8 am to 6 pm) were 22.2 $^{\circ}$C and 21.2 $^{\circ}$C, respectively. The HVAC system was shut down during unoccupied hours.

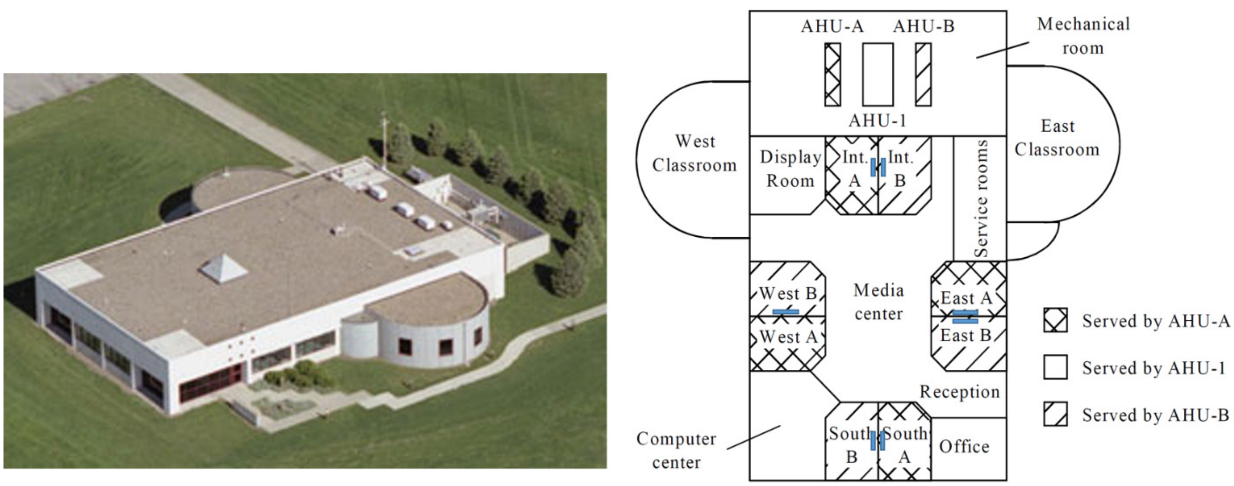


Figure **12**Energy resource station at Iowa Energy Center[58]

We follow the exact same experimental settings, including the collection of operational features, the derivation of the meta-features, the training data selection and cross-validation. Again, the validation is conducted on both a single day test and one week test. Since the measurement data is collected between August and September, while the BEMR is built based on July, this could be viewed as an extrapolation test. The performance rankings of the six forecasting model along with the predicted rankings from BEMRare provided in Table 7.

Table 7Performance Rankings (T) of the Six Forecasting Models and the Predicted Rankings from BEMR (B) on Single Day and One Week Tests

Model	Mon		Tue		Wed		Thu		Fri		Sat		Sun		One Week	
	T	B	T	B	T	B	T	B	T	B	T	B	T	B	T	B
Kriging	1	1	2	2	2	2	2	1	1	1	3	2	3	4	1	1
SVR	5	5	5	5	6	5	5	4	3	4	2	3	1	2	6	6
RBF	4	3	4	3	5	6	4	5	5	6	6	6	6	5	5	5
MARS	2	6	6	6	4	4	3	3	6	5	1	1	3	3	1	2
ANN	6	4	3	4	1	1	6	6	2	2	5	5	2	1	4	3
PR	2	2	1	1	3	3	1	2	3	3	3	4	5	6	3	4

The statistics show that 2 out of 8 test cases (Thursday and Sunday) are assigned a sub-optimal model, while the assigned models are both ranked second according to the ground truth, which implies the BEMR generally provides reliable recommendations. In conclusion, the validation experiment results on the real building indicate that the proposed BEMR is able to assist real building energy forecasting tasks with reliable and high quality solutions.

4 DISCUSSION,CONCLUSION AND FUTURE WORK

This paper is motivated to develop a computationally efficient data-driven approach to quickly identify appropriate algorithmsfor building energy load forecasting.We propose a recommendation system for short-term building forecasting model selection based on a meta-learning technique. This is an extensively studied automatic learning algorithm applied to meta-data in machine learning experiments. We propose various meta-features which characterize the building energy data: building electricity load time series features, building operational features and physical features. An Artificial Neural Network is applied to model the relationship between the meta-features and the ranking of each model derived from the performance on forecasting. In addition, due to the high dimensionality of the proposed meta-features, an advanced feature reduction technique, Singular Value Decomposition, is applied on the meta-feature space to improve the meta-learning performance and reduce computational cost. The resulting high spearman's ranking correlation coefficient and success rate on the two test cases: single day and one week, and the extrapolation test, indicate the successful implementation of the recommendation system.

To demonstrate the applicability of the proposed recommendation system, 48 benchmark buildingshave been tested, including 8 types of typical buildingslocated across 6 climate zones coveringa wide range of building profiles. One real building is used to validate the system for assessment of the applicability and extensibility to real problems. To evaluate the forecasting capabilityof the proposed framework, we have also implemented various popular data-driven forecasting methods in the literature, including Kriging, SVR,RBF, MARS, ANN and PR. Regarding the practical advantages of this framework and its combination with energy supplies in the domain of building energy and power systems, the proposed recommendation system can be used to facilitate the development of a building energy expert system for real-time building operations management, decision making and support.Comparing this technique to the traditional approach, it is concluded that the meta-learning approach can achieve both high prediction accuracy and high computational efficiency on various genres of building forecasting problems.It augmentsthe traditional trial-and-error meta-modeling method in that itenablesan automated and optimizedmodeling process which requires littleexpert involvement and minimizes excessive computations. Based on past experience, the recommendation system emulates the human's decision-makingability, which makes reasonable decisions and efficient calculations to solve complex problems. Specifically, it consists of a two stagelearning process: the knowledge base is first constructed, which accumulates facts and rules about the problem domain, andthen an inference engine is builtto apply the rules from the known facts anddeduce new facts. This work provides practical guidelines in the design, development, implementation, and testing of a forecasting recommendation system for various short-term building energy forecasting problems. Specifically,it can help non-experts with forecasting model selection.Due to these theoretical contributions and advantages, we recommend its use to facilitate everyday building energy industrial applications and operationsto reduce the cost and improve modeling and operation efficiency.

In summary, the originality of this paper isthree-fold:

- The first contribution is the implementation of a two-stage meta-learning framework on varioustime-series problems in the domain of building energy modeling.
- The secondcontribution stems from the proposed generalized automatic meta-learning based expert system which requires little human involvementto support forecasting model recommendation.
- To the best of our knowledge, this is the first recommendation system motivated from the machine learning domain for shortterm building forecasting based on various meta-features derived from both of building data-characteristics and physical-characteristic features.

We acknowledge that conducting our analysis in the scope of STLF is a limitation of this study. However,the proposed approach adequately demonstrates the applicability of the recommendation system on energy forecasting for various types of buildings across different climate zones. We envision that the STLF framework is viably transformable to MTLF and LTLF by adjusting the operational features and meta-features, and we reserve this for our future work.

APPENDIX DATA-DRIVEN MODELING TECHNIQUES

The data-driven modeling techniques build models solely on historical data which is represented by the time-delay variables, e.g., temperature, humidity, and past energy consumption data, that form the feature vectors. This makes the forecasting process more adaptive to different types of buildings and reduces human involvement for model adjustments [6].Six selected data-driven modeling techniques, including four of statistical modeling methods, Kriging, RBF, MARS and PR, and two machine learning methods, SVR and ANN, are reviewed in this section. Notice in each method thatwe are building a model of the building energy consumption based on the building'soperational features.

Kriging

Kriging (also known as Gaussian process regression) is an interpolation method that assumes the simulation output may be modeled by a Gaussian process. It gives the best linear unbiased prediction of simulation output not yet observed. It generates the prediction in the form of a combination of a global model with local random noise:

$$\boldsymbol{y}(\boldsymbol{x}) = \boldsymbol{f}(\boldsymbol{x})\boldsymbol{\beta} + \boldsymbol{Z}(\boldsymbol{x}), \tag{1}$$

where *x* is the input vector, $\beta$is the weight vector, and *Z(x)* is a stochastic process with zero mean and stationary covariance of

$$\boldsymbol{cov}\left[\mathbf{Z}(\boldsymbol{x_i}), \mathbf{Z}(\boldsymbol{x_j})\right] = \boldsymbol{\sigma^2}\boldsymbol{R}\left(\boldsymbol{x_i}, \boldsymbol{x_j}\right), \tag{2}$$

where σ^2 is the process variance, $R(x_i, x_j)$ is an *n* by *n* correlation matrix where *n* is the sample size of the training data. R is usually depicted by a Gaussian correlation function, $exp(-\theta(x_i - x_j)^2)$ with parameterθ.

Support Vector Regression

Support Vector Regression (SVR) is analogous to support vector classification, which attempts to maximize the distance between two classes of data by selecting two hyperplanes to optimally separate the training data. The mathematical form of SVR is:

$$\boldsymbol{f(x)} = \langle \boldsymbol{\omega} \cdot \boldsymbol{x} \rangle + \boldsymbol{b}, \tag{3}$$

where ω is the norm vector to the hyperplane and $b/\|\omega\|$determines the offset of the hyperplane from the origin. The goal is to find a hyperplane that separates the data points optimally without error and separates the closest points with the hyperplane as far as possible. Thus, it can be constructed as an optimization problem:

$$min 1/2|\omega|^2$$

$$\text{s.t.}\begin{cases} \boldsymbol{y_i} - \langle \boldsymbol{\omega} \cdot \boldsymbol{x_i} \rangle - \boldsymbol{b} \leq \boldsymbol{\varepsilon} \\ \langle \boldsymbol{\omega} \cdot \boldsymbol{x_i} \rangle + \boldsymbol{b} - \boldsymbol{y_i} \leq \boldsymbol{\varepsilon}' \end{cases} \tag{4}$$

According to the duality principle, the nonlinear regression problem is given by:

$$\boldsymbol{f(x)} = \sum_{\boldsymbol{i=1}}^{\boldsymbol{m}} (\boldsymbol{\alpha_i^*} - \boldsymbol{\alpha_i})\, \boldsymbol{k}\langle \boldsymbol{x_i} \cdot \boldsymbol{x_j} \rangle + \boldsymbol{b}. \tag{5}$$

where α_i^* and α_i are two introduced dual variables, and $k\langle x_i \cdot x_j \rangle$ is a kernel function, e.g. Gaussian kernel.

Radial Basis Function

Radial Basis Function (RBF) is used to develop interpolation on scattered multivariate data. A RBF is a linear combination of a real-valued radially symmetric function,$\emptyset(x)$, based on distance from the origin,

$$f(x) = \sum_{i=1}^{n} \theta_i\, \emptyset(\|x - x_i\|). \tag{6}$$

where θ_i is the unknown interpolation coefficient determined by the least-squares method, n is the number of sampling points and $\|x - x_i\|$ is the Euclidean norm of the radial distance from design point x to the sampling point x_i.

Multivariate Adaptive Regression Splines

Multivariate Adaptive Regression Splines (MARS) is a form of regression analysis introduced by Friedman(1991). A set of basis functions, defined as constant, hinge function, or the product of two or more hinge functions, are combined in the weighted sum form, to be the approximation of the response function. A MARS model is built with generalized cross validation regularization in a forward/backward iterative process. The general model of MARS can be written as:

$$f(x) = \gamma_0 + \sum_{i=1}^{m} \gamma_i h_i(x), \tag{7}$$

where γ_i is the constant coefficient of the combination whose value is jointly adjusted to give the best fit to the data, and the basis function h_i, can be represented as:

$$h_i(x) = \prod_{k=1}^{K_m} [s_{k,m} \cdot (x_{v(k,m)} - t_{k,m})]_+^q. \tag{8}$$

where K_m is the number of splits given to the m^{th} basis function, $s_{k,m}$=± 1 indicates the right/left sense of the associated step function, $v(k,m)$ is the label of the variable, and $t_{k,m}$ represents values (knot locations) of the corresponding variables. The superscript q and subscript + indicate the truncated power functions with polynomials of lower order than q.

Artificial Neural Network

Artificial Neural Network (ANN) [60] is a computational model inspired by an animal's central nervous system. It is apt at solving problems with complicated structures. Due to its promising results in numerous fields, ANN has been extensively applied in stochastic simulation modeling (Fonseca, Navaresse, & Moynihan, 2003; Nasereddin & Mollaghasemi,1999). An ANN model typically consists of three separate layers: the input layer, the hidden layer(s), and the output layer. The neurons across different layers are interconnected to transmit and deduce information. A typical ANN with three layers and one single output neuron has the following mathematical form:

$$f(x) = \sum_{j=1}^{J} \omega_j \delta(\sum_{i=1}^{I} w_{ij} \delta(x_i) + \alpha_j) + \beta + \varepsilon. \tag{9}$$

where x is a k-dimensional vector, the input unit represents the raw information that is fed into the network, $\delta(\cdot)$ is the user defined transfer function, w_{ij} is the weight factor on the connection between the i^{th} input neuron and the j^{th} hidden neuron, α_j is the bias in the j^{th} hidden neuron, ω_j is the weight on connection between the j^{th} hidden neuron and the output neuron, β is the bias of the output neuron, ε is a random error with a mean of 0, and I and J are the number of input neurons and hidden neurons. In supervised learning, the output unit is trained to simulate the underlying structure of the input signals and response. The trained structure is depicted by several parameters, the weights on each connection, the biases, the number of hidden layers, the transfer functions, and the number of hidden nodes in each hidden layer.

Polynomial Regression

Polynomial Regression (PR) is a variation of linear regression in which an n^{th} order polynomial is modeled to formulate the relationship between the independent variable x and the dependent variable y. PR models have been applied to various engineering domains such as mechanical, medical and industrial (Barker et al., 2001;Greenland, 1995; Shaw et al., 2006). A second-order polynomial model can be expressed as:

$$f(x) = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_{ii} {x_i}^2 + \sum_i \sum_j \beta_{ij} x_i \, x_j + \epsilon \tag{10}$$

where β is the constant coefficient, k is the number of variables, ε is an unobserved random error with zero mean, PR models are usually fit using the least squares method.
Extensive applications on forecasting using the reviewed techniques canbe found in [5], [18], [65], [66].

Acknowledgement

This research was partially supported by funds from the National Science Foundation award under grant number CNS-1239257 and from the United States Transportation Command(USTRANSCOM) in concert with the Air Force Institute of Technology (AFIT) under an ongoing Memorandum of Agreement. The U.S. Government is authorized to reproduce and distribute for governmental purposes notwithstanding any copyright annotation of the work by the author(s). The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of USTRANSCOM, AFIT, the Department of Defense, or the U.S. Government.

References

[1] Architecture 2030, “2030 Challenge for Products: Critical Points,” 2011.

[2] X. Li, J. Wen, and E.-W. Bai, “Developing a whole building cooling energy forecasting model for on-line operation optimization using proactive system identification,” *Appl. Energy*, vol. 164, no. 2016, pp. 69–88, 2016.

[3] X. Li and J. Wen, “Review of building energy modeling for control and operation,” *Renew. Sustain. Energy Rev.*, vol. 37, pp. 517–537, 2014.

[4] B. Eisenhower, Z. O’Neill, S. Narayanan, V. a. Fonoberov, and I. Mezić, “A methodology for meta-model based optimization in building energy models,” *Energy Build.*, vol. 47, pp. 292–301, Apr. 2012.

[5] L. Yu, S. Wang, and K. K. Lai, “A neural-network-based nonlinear metamodeling approach to financial time series forecasting,” *Appl. Soft Comput.*, vol. 9, no. 2, pp. 563–574, Mar. 2009.

[6] D. Solomon, R. Winter, A. Boulanger, R. Anderson, and L. Wu, “Forecasting Energy Demand in Large Commercial Buildings Using Support Vector Machine Regression,” 2000.

[7] J. Crespo Cuaresma, J. Hlouskova, S. Kossmeier, and M. Obersteiner, “Forecasting electricity spot-prices using linear univariate time-series models,” *Appl. Energy*, vol. 77, no. 1, pp. 87–106, Jan. 2004.

[8] W.-C. Hong, “Electric load forecasting by seasonal recurrent SVR (support vector regression) with chaotic artificial bee colony algorithm,” *Energy*, vol. 36, no. 9, pp. 5568–5578, Sep. 2011.

[9] T. Salsbury and R. Diamond, “Performance Validation and Energy Analysis of HVAC Systems using Simulation 1 Introduction 2 Simulation-Based Validation Methodology,” *Indoor Environ.*, vol. 25, pp. 1–20, 1996.

[10] M. Hu, “A data-driven feed-forward decision framework for building clusters operation under uncertainty,” *Appl. Energy*, vol. 141, no. 2015, pp. 229–237, 2015.

[11] M. Hu and H. Cho, “A probability constrained multi-objective optimization model for CCHP system operation decision support,” *Appl. Energy*, vol. 116, pp. 230–242, 2014.

[12] T. Hong, “Short Term Electric Load Forecasting,” 2010.

[13] H . Lee Willis, *Power Distribution Planning Reference Book*. CRC press, 2004.

[14] H. Wang, P. Xu, X. Lu, and D. Yuan, “Methodology of comprehensive building energy performance diagnosis for large commercial buildings at multiple levels,” *Appl. Energy*, vol. 169, no. 2016, pp. 14–27, 2016.

[15] M. T. Hagan and S. M. Behr, “The Time Series Approach to Short Term Load Forecasting,” vol. PWRS-2, no. 3, pp. 785–791, 1987.

[16] R. J. Hyndman, A. B. Koehler, R. D. Snyder, and S. Grose, “A state space framework for automatic forecasting using exponential smoothing methods,” *Int. J. Forecast.*, vol. 18, no. 3, pp. 439–454, 2002.

[17] S. Baldi, S. Yuan, P. Endel, and O. Holub, “Dual estimation: Constructing building energy models from data sampled at low rate,” *Appl. Energy*, vol. 169, pp. 81–92, 2016.

[18] L. E. Mavromatidis, A. Bykalyuk, and H. Lequay, “Development of polynomial regression models for composite dynamic envelopes’ thermal performance forecasting,” *Appl. Energy*, vol. 104, pp. 379–391, 2013.

[19] H. S. Hippert, C. E. Pedreira, and R. C. Souza, “Neural networks for short-term load forecasting: a review and evaluation,” *IEEE Trans. Power Syst.*, vol. 16, no. 1, pp. 44–55, 2001.

[20] C. R. Touretzky and R. Patil, “Building-level power demand forecasting framework using building specific inputs: Development and applications,” *Appl. Energy*, vol. 147, pp. 466–477, 2015.

[21] A. D. Papalexopoulos and T. C. Hesterberg, “A regression-based approach to short-term system load forecasting,” *IEEE Trans. Power Syst.*, vol. 5, pp. 1535–1547, 1990.

[22] J. W. Taylor and P. E. McSharry, “Short-term load forecasting methods: An evaluation based on European data,” *IEEE Trans. Power Syst.*, vol. 22, pp. 2213–2219, 2007.

[23] N. K. Ahmed, A. F. Atiya, N. El Gayar, and H. El-Shishiny, “An Empirical Comparison of Machine Learning Models for Time Series Forecasting,” *Econom. Rev.*, vol. 29, no. 5–6, pp. 594–

621, Aug. 2010.

[24] S. Chirarattananon and J. Taveekun, "An OTTV-based energy estimation model for commercial buildings in Thailand," *Energy Build.*, vol. 36, no. 7, pp. 680–689, 2004.

[25] F. W. H. Yik, J. Burnett, and I. Prescott, "Predicting air-conditioning energy consumption of a group of buildings using different heat rejection methods," *Energy Build.*, vol. 33, no. 2, pp. 151–166, 2001.

[26] L. Tang, L. Yu, and K. He, "A novel data-characteristic-driven modeling methodology for nuclear energy consumption forecasting," *Appl. Energy*, vol. 128, no. 2014, pp. 1–14, 2014.

[27] J. S. Armstrong, "Forecasting by Extrapolation : Conclusions from Twenty-five Years of Research," vol. 14, no. 6, 1984.

[28] X. Lü, T. Lu, C. J. Kibert, and M. Viljanen, "Modeling and forecasting energy consumption for heterogeneous buildings using a physical–statistical approach," *Appl. Energy*, vol. 144, no. 2015, pp. 261–275, 2015.

[29] M. Matijaš, "Electric Load Forecasting using Multivariate Meta-learning," 2013.

[30] B. F. De Souza, A. De Carvalho, and C. Soares, "Metalearning for Gene Expression Data Classification," *2008 Eighth Int. Conf. Hybrid Intell. Syst.*, pp. 441–446, Sep. 2008.

[31] Z. Lan, J. Gu, Z. Zheng, R. Thakur, and S. Coghlan, "A study of dynamic meta-learning for failure prediction in large-scale systems," *J. Parallel Distrib. Comput.*, vol. 70, no. 6, pp. 630–643, Jun. 2010.

[32] S. Zhou, K. K. Lai, and J. Yen, "A dynamic meta-learning rate-based model for gold market forecasting," *Expert Syst. Appl.*, vol. 39, no. 6, pp. 6168–6173, May 2012.

[33] Marin Matijaš, "Electric Load Forecasting Using Support Vector.pdf," 2013.

[34] F. Collopy and J. S. Armstrong, "Rule-Based Forecasting: Development and Validation of an Expert Systems Approach to Combining Time Series Extrapolations," *Management Science*, vol. 38, no. 10. pp. 1394–1414, 1992.

[35] R. B. C. Prudêncio and T. B. Ludermir, "Meta-learning approaches to selecting time series models," *Neurocomputing*, vol. 61, pp. 121–137, Oct. 2004.

[36] X. Wang, K. Smith-Miles, and R. Hyndman, "Rule induction for forecasting method selection: Meta-learning the characteristics of univariate time series," *Neurocomputing*, vol. 72, pp. 2581–2594, 2009.

[37] C. Lemke and B. Gabrys, "Meta-learning for time series forecasting and forecast combination," *Neurocomputing*, vol. 73, no. 10–12, pp. 2006–2016, Jun. 2010.

[38] M. Matijaš, J. a. K. Suykens, and S. Krajcar, "Load forecasting using a multivariate meta-learning system," *Expert Syst. Appl.*, vol. 40, no. 11, pp. 1–11, 2013.

[39] C. Cui, M. Hu, J. D. Weir, and T. Wu, "A Recommendation System for Meta-modeling: A Meta-learning based Approach," *Expert Syst. Appl.*, 2015.

[40] G. G. Wang and S. Shan, "Review of Metamodeling Techniques in Support of Engineering Design Optimization," *J. Mech. Des.*, vol. 129, no. 4, p. 370, 2007.

[41] Anna Ściążko, "Surrogat modeling techniques applied to energy systems," University of Iceland & University of Akureyri, 2011.

[42] H. X. Zhao and F. Magoulès, “A review on the prediction of building energy consumption,” *Renew. Sustain. Energy Rev.*, vol. 16, no. 6, pp. 3586–3592, 2012.

[43] M. Deru, K. Field, D. Studer, K. Benne, B. Griffith, P. Torcellini, B. Liu, M. Halverson, D. Winiarski, and M. Rosenberg, “U.S. Department of Energy Commercial Reference Building Models of the National Building Stock,” 2011.

[44] ASHRAE., “Energy Efficient Design of New Buildings Except Low-Rise Residential Buildings. ANSI/ASHRAE/IESNA Standard 90.1 -2004.” American Society of Heating, Refrigeration, and Air-Conditioning Engineers, Atlanta, GA, 2004.

[45] S. Wilcox and W. Marion, “Users manual for TMY3 data sets,” 2008.

[46] D. C. Montgomery, E. a Peck, and G. G. Vining, *Introduction to Linear Regression Analysis*. John Wiley & Sons, 2012.

[47] X. Li, C. Sun, and D. Gong, “Application of support vector machine and similar day method for load forecasting,” in *Advances in Natural Computation*, Springer, 2005, pp. 602–609.

[48] J. Racine, “Consistent cross-validatory model-selection for dependent data : hv -block cross-validation,” *J. Econom.*, vol. 99, pp. 39–61, 2000.

[49] D. Kugiumtzis, “Surrogate data test for nonlinearity including nonmonotonic transforms,” *Phys. Rev. E - Stat. Physics, Plasmas, Fluids, Relat. Interdiscip. Top.*, vol. 62, no. 1, pp. 25–28, 2000.

[50] D. J. Fonseca, D. O. Navaresse, and G. P. Moynihan, “Simulation metamodeling through artificial neural networks,” *Eng. Appl. Artif. Intell.*, vol. 16, pp. 177–183, 2003.

[51] L. Yu, S. Wang, and K. Keung, “Forecasting crude oil price with an EMD-based neural network ensemble learning paradigm,” vol. 30, pp. 2623–2635, 2008.

[52] G. Hughes, “On the mean accuracy of statistical pattern recognizers,” *IEEE Trans. Inf. Theory*, vol. 14, 1968.

[53] C. Eckart and G. Young, “The approximation of one matrix by another of lower rank,” *Psychometrika*, vol. 1, pp. 211–218, 1936.

[54] H. R. Neave and P. L. Worthington, *Distribution-free tests*. London: Routledge, 1989.

[55] R. K. S. et al. Crawley, Drury B., Linda K. Lawrie, Frederick C. Winkelmann, Walter F. Buhl, Y. Joe Huang, Curtis O. Pedersen, “EnergyPlus: creating a new-generation building energy simulation program,” *Energy Build.*, vol. 33, no. 4, pp. 319–331, 2001.

[56] C. Cui, T. Wu, M. Hu, J. D. Weir, and X. Chu, “Accuracy vs. robustness: bi-criteria optimizaed ensemble of metamodels,” *Proc. 2014 Winter Simul. Conf.*, pp. 616–627, 2014.

[57] J Wen and S. Li, “Tools for evaluating fault detection and diagnostic methods for air-handling units,” ASHRAE, 2012.

[58] B. A. Price and T. F. Smith, “Development and validation of adaptive optimal operation methodology for building HVAC systems,” 2003.

[59] J. H. Friedman, “Multivariate Adaptive Regression Splines,” *Ann. Stat.*, vol. 19, pp. 1–67, 1991.

[60] F. Rosenblatt, “The perceptron: a probabilistic model for information storage and organization in the brain.,” *Psychol. Rev.*, vol. 65, no. 6, pp. 386–408, 1958.

[61] M. Nasereddin and M. Mollaghasemi, “The development of a methodology for the use of neural networks and simulation modeling in system design,” in *1999 Winter Simulation Conference*

(WSC'99) - Volume 1, 1999, pp. 537–542.

[62] S. Greenland, "Dose-response and trend analysis in epidemiology: alternatives to categorical analysis.," *Epidemiology*, vol. 6, pp. 356–365, 1995.

[63] P. A. Barker, F. A. Street-Perrott, M. J. Leng, P. B. Greenwood, D. L. Swain, R. A. Perrott, R. J. Telford, and K. J. Ficken, "A 14,000-year oxygen isotope record from diatom silica in two alpine lakes on Mt. Kenya.," *Science*, vol. 292, pp. 2307–2310, 2001.

[64] P. Shaw, D. Greenstein, J. Lerch, L. Clasen, R. Lenroot, N. Gogtay, A. Evans, J. Rapoport, and J. Giedd, "Intellectual ability and cortical development in children and adolescents.," *Nature*, vol. 440, pp. 676–679, 2006.

[65] S. K. Sahu, "A Bayesian Kriged-Kalman model for short-term fore- casting of air pollution levels," *Appl. Stat.*, vol. 54, pp. 223–244, 2005.

[66] W. Shen, X. Guo, C. Wu, and D. Wu, "Forecasting stock indices using radial basis function neural networks optimized by artificial fish swarm algorithm," *Knowledge-Based Syst.*, vol. 24, no. 3, pp. 378–385, 2011.